Bayesian estimates of parameter variability in the $k - \varepsilon$ turbulence model

W.N. Edeling$^{a,b}$, P. Cinnella$^{a}$, R.P. Dwight$^{b,*}$, H. Bijl$^{b}$

$^a$ENSAIM ParisTech, DynFluid laboratory, 151 Boulevard de l’Hospital, 75013 Paris, France
$^b$Delft University of Technology, Faculty of Aerospace Engineering, Kluyverweg 2, Delft, The Netherlands

Abstract

In this paper we are concerned with obtaining estimates for the error in Reynolds-Averaged Navier-Stokes (RANS) simulations based on the Launder-Sharma $k - \varepsilon$ turbulence closure model, for a limited class of flows. In particular we search for estimates grounded in uncertainties in the space of model closure coefficients, for wall-bounded flows at a variety of favourable and adverse pressure gradients. In order to estimate the spread of closure coefficients which reproduces these flows accurately, we perform 13 separate Bayesian calibrations – each at a different pressure gradient – using measured boundary-layer velocity profiles, and a statistical model containing a multiplicative model inadequacy term in the solution space. The results are 13 joint posterior distributions over coefficients and hyper-parameters. To summarize this information we compute Highest Posterior-Density (HPD) intervals, and subsequently represent the total solution uncertainty with a probability-box (p-box). This p-box represents both parameter variability across flows, and epistemic uncertainty within each calibration. A prediction of a new boundary-layer flow is made with uncertainty bars generated from this uncertainty information, and the resulting error estimate is shown to be consistent with measurement data.

Keywords: Bayesian calibration, parameter variability, model inadequacy, RANS turbulence model, global sensitivity analysis

1. Introduction

Computational Fluid Dynamics (CFD) and Reynolds-averaged Navier-Stokes (RANS) simulations in particular form an important part of the analysis and design methods used in industry. These simulations are typically based on a deterministic set of input variables and model coefficients. However real-world

*Corresponding author

Email addresses: W.N.Edeling@tudelft.nl (W.N. Edeling), P.Cinnella@ensam.eu (P. Cinnella), R.P.Dwight@tudelft.nl (R.P. Dwight), H.Bijl@tudelft.nl (H. Bijl)
flow problems are subject to numerous uncertainties, e.g. imprecisely known parameters, initial- and boundary conditions. For input uncertainties described as probability density functions (pdfs), established methods exist for determining the corresponding output uncertainty [6, 45, 7]. Furthermore, numerical predictions are affected by numerical discretization errors and approximate physical models (turbulence models in RANS). The former may be estimated and controlled by means of mesh refinement (e.g. Ref. 8), but no analogue exists for the latter. This error, which we call model inadequacy in the following, is therefore the only major source of simulation error that remains difficult to estimate. It is therefore the bottleneck in the trustworthiness of RANS simulations. This work is an attempt to construct an estimate of model inadequacy in RANS for a limited set of flows, and for a single turbulence closure model.

Within the framework of RANS, many turbulence models are available, see e.g. Ref. 44 for a review. There is general agreement that no universally-”best” RANS turbulence closure model is currently known: the accuracy of models is problem-dependent [47]. Moreover, each turbulence model uses a number of closure coefficients which are classically determined by calibration against a database of fundamental flows [35]. Model performance may strongly depend on these values, which are often adjusted to improve model accuracy for a given set of problems, or for a specific flow code. They are almost always assumed to be constant in space and time. For a given model there is sometimes no consensus on the best values for these coefficients, and often intervals are proposed in the literature [31].

Our approach is to represent model inadequacy by uncertainty in these coefficients. Summarized we proceed as follows: (1) we define the class of flows for which we wish to estimate the error, in this article turbulent boundary-layers for a range of pressure gradients. (2) We collect experimental data for a number of flows of this class. (3) We use Bayesian model updating to calibrate the closure coefficients against each flow in this data-set, resulting in posterior distributions on the coefficients for each flow [10]. (4) We summarize the large amount of posterior information using highest posterior-density (HPD) intervals. This summary gives intervals on the coefficients which represent both the spread of coefficients within the flow-class, as well as the ability of the calibration to provide information about the values these coefficients should take in each flow. (5) For a new flow of the class, for which there might be no experimental data, we then perform a simulation using the model with the specified coefficient uncertainties. The resulting interval on the model output is our probabilistic estimate of the true flow.

Representing model inadequacy by uncertainty in closure coefficients is reasonable since the coefficients are empirical: they must be seen as ”tuning parameters” associated to the model and, in general, they are not expected to be flow-independent. Furthermore each coefficient is involved in an approximation of the underlying physics, and therefore is closely related to some component of the model inadequacy. Finally an error estimate based on coefficient uncertainty has the virtue of being geometry-independent – that is we do not need to assume a particular flow topology to apply the estimate. We do not claim that
it is possible to approximate all turbulence model inadequacy in this way. The method does rely on being able to approximate most of it, and we demonstrate that this is possible for the limited class of flows we consider.

The key step in the method is the calibration of the coefficients. For the calibration phase we follow the work of Cheung et al. [3], in which a Bayesian approach was applied to the calibration of the Spalart-Allmaras [37] turbulence model, taking into account measurement error [21]. In that work, for a given statistical model, the coefficients were calibrated once on all the available measured velocity profiles and wall-shear stress components. Model inadequacy was treated with a multiplicative term parameterized in the wall-normal direction with a Gaussian process, following the framework of Kennedy and O’Hagan [15]. In the present work, we perform an analysis by performing separate calibrations on multiple flows in our class, using the $k-\varepsilon$ model, with Launder-Sharma damping functions [19]. Using uniform priors and calibrating against a large, accurate data-set containing boundary-layer profiles at different pressure gradients, results in informative coefficient posteriors for each flow. The multiplicative model inadequacy term is retained to capture the part of the error which cannot be captured by the closure coefficients alone.

We choose the pressure gradient as the independent variable in our flow class because it is known to have a large impact on the performance of $k-\varepsilon$ model [17, 32, 42]. Approaching this problem in a Bayesian context allows us to estimate how much this deficiency can be reduced by choice of closure coefficients alone, and how much the coefficients have to vary to match measurements at all pressure-gradients. The spread of coefficients is an indication of flow-independence of the model, and we expect better models to have smaller spreads.

The paper is laid out as follows: we briefly outline the framework of Bayesian data analysis in Section 2, and the $k-\varepsilon$ model in Section 3. Section 4 describes the experimental data used for the calibration, and Section 5 describes our calibration framework, in particular the statistical model and priors. The results, including verification, HPD analysis of calibration posteriors, and prediction using the obtained coefficient uncertainties are described in Section 6. Specifically, our confidence interval estimate for error due to turbulence modelling inadequacy is given in Section 6.7. Finally, Section 7 summarizes the main findings and provides guidelines for future research.

2. General Bayesian Data Analysis

By Bayesian data analysis, we mean practical methods for making inferences from data using joint probability models for quantities we observe and for quantities we wish to learn about [10]. Bayesian data analysis is often applied to problems in uncertainty quantification, see e.g. Refs. 4, 27, 9. One type of uncertainty, namely aleatoric uncertainty, arises through natural random variations of the process. This type of uncertainty is irreducible (intrinsic to the system), such that more data or better models will not reduce it. Epistemic uncertainty on the other hand, arises from a lack of knowledge about the model,
e.g. unknown model parameters or mathematical form, and can in principle be reduced. The Bayesian framework represents epistemic uncertainty using probability, which is often used to represent only aleatory uncertainty. Hence all sources of uncertainty are probabilistic, leading to a unified treatment exploiting the tools of uncertainty quantification and Bayes’ theorem. The general process for Bayesian data analysis consists of four steps: 1) define a joint-probability distribution for both the observed and unobserved quantities in the problem, 2) calibrate the model against observations, 3) validate the model, and finally 4) use the calibrated model to make predictions.

Inferences are made for two kinds of unobserved quantities, namely

1. **Parameters** that govern the model, which we denote by the column vector \( \mathbf{\theta} \), and are treated as random variables.
2. **Future predictions** of the model. If we let \( \mathbf{z} = [z_1, z_2, \ldots, z_n] \) denote the observed data, then the currently unknown (but possibly observable) future predictions are denoted by \( \tilde{\mathbf{z}} \).

In addition to parameters subject to calibration, we have a class of *explanatory* variables \( \mathbf{t} \). These are the variables that we do not bother to model as random, e.g. the pressure gradients in our boundary layer calibrations, but do affect the model predictions.

In short, the aim of Bayesian data analysis is to draw conclusions about \( \mathbf{\theta} \) (calibration) through the conditional posterior distribution \( p(\mathbf{\theta} | \mathbf{z}) \), or about \( \tilde{\mathbf{z}} \) (prediction) through \( p(\tilde{\mathbf{z}} | \mathbf{z}) \). Here, we let \( p(\cdot) \) denote a probability distribution. We can achieve this via the application of *Bayes’ theorem*

\[
p(\mathbf{\theta} | \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{\theta}) p(\mathbf{\theta})}{p(\mathbf{z})}, \quad (1)
\]

where the law of total probability states that \( p(\mathbf{z}) = \int p(\mathbf{z} | \mathbf{\theta}) p(\mathbf{\theta}) d\mathbf{\theta} \). Since this denominator does not depend upon \( \mathbf{\theta} \), it is often omitted to yield the unnormalized version of (1),

\[
p(\mathbf{\theta} | \mathbf{z}) \propto p(\mathbf{z} | \mathbf{\theta}) p(\mathbf{\theta}). \quad (2)
\]

The term \( p(\mathbf{z} | \mathbf{\theta}) \), i.e. the distribution of the data given the parameters is called the *likelihood function*, and it provides the means for updating the model once more data becomes available. The term \( p(\mathbf{\theta}) \) is the *prior distribution* of \( \mathbf{\theta} \), i.e. it represents what we know about the parameters before the data became available.

The posterior predictive distribution conditional on the observed \( \mathbf{z} \) can be written as

\[
p(\tilde{\mathbf{z}} | \mathbf{z}) = \int p(\tilde{\mathbf{z}}, \mathbf{\theta} | \mathbf{z}) d\mathbf{\theta} = \int p(\tilde{\mathbf{z}} | \mathbf{\theta}, \mathbf{z}) p(\mathbf{\theta} | \mathbf{z}) d\mathbf{\theta} = \\
\int p(\tilde{\mathbf{z}} | \mathbf{\theta}) p(\mathbf{\theta} | \mathbf{z}) d\mathbf{\theta}. \quad (3)
\]
The last step follows because $\tilde{z}$ and $z$ are assumed to be conditionally independent given $\theta$, i.e. $p(\tilde{z} \mid z, \theta) = p(\tilde{z} \mid \theta)$.

In general the relationship between $\theta$ and $z$ involves evaluation of a numerical computer code. Therefore the posterior distribution of coefficients, $p(\theta \mid z)$, and integrals such as (3) cannot be evaluated analytically. Evaluation of $p(\theta \mid z)$ at a single value of $\theta$ requires at least one flow-solve, so for moderate-dimensional $\theta$ brute-force sampling of the posterior is computationally intractable. Markov-chain Monte-Carlo (MCMC) methods \cite{23, 11} are used to provide samples $\theta^j, j = 1, 2, \cdots, J$ from $p(\theta \mid z)$ at a more acceptable cost. The form of the right-hand side of (3) then suggests using these samples to draw samples $\tilde{z}^j$ from the conditional distribution $p(\tilde{z} \mid \theta)$, at which point the integral can be approximated numerically by standard Monte-Carlo.

In the case of CFD applications, sampling $\theta$ could be much too expensive such that even MCMC sampling requires an unacceptably high overall computational time. This problem may be alleviated by replacing the CFD model with an approximated inexpensive analytical model, a so-called surrogate model, such as polynomial interpolation or Kriging (see e.g. Ref. 7). This is not used in the present work since the boundary layer code used for computing the numerical solution is cheap enough to be directly coupled to MCMC.

Note the result of the analysis depends on the chosen stochastic model (which defines the joint pdf $p(z, \theta)$), hence the necessity for a post-calibration validation step. A way to reduce the bias introduced by the choice of a single stochastic model is to apply Bayesian model averaging, in which a Bayesian analysis is performed using multiple stochastic models $M_i$ from a set of competing model classes $M = \{M_1, \cdots, M_m\}$. This avoids having to choose one model, by creating a weighted average model constructed from the models in the set $M$. The weights of the models are unknowns to be calibrated, and they can be interpreted as the level of evidence for a given model \cite{12}.

3. The $k - \varepsilon$ turbulence model

The general simulation approach considered in this paper is the solution of the RANS equations for turbulent boundary layers, supplemented by a turbulence model. RANS equations remain up to now the most advanced and yet computationally acceptable simulation tool for engineering practice, since more advanced strategies, like Large Eddy Simulation (see e.g. Ref. 33) are yet too expensive for high-Reynolds flows typically encountered in practical applications. Under the assumption of incompressibility, the governing equations for a boundary-layer flow are given by

$$\frac{\partial \bar{u}_1}{\partial t} + \bar{u}_1 \frac{\partial \bar{u}_1}{\partial x_1} + \bar{u}_2 \frac{\partial \bar{u}_1}{\partial x_2} = -\frac{1}{\rho} \frac{\partial \bar{p}}{\partial x_2} + \frac{\partial}{\partial x_2} \left[ (\nu + \nu_T) \frac{\partial \bar{u}_1}{\partial x_2} \right],$$

where $\rho$ is the constant density, $\bar{u}_i$ is the mean velocity in $x_i$ direction and $\nu$ is the kinematic viscosity. The eddy viscosity $\nu_T$ is meant to represent the effect
of turbulent fluctuations on the mean flow, and is calculated here through the $k - \varepsilon$ turbulence model:

$$\nu_T = C_\mu f_\mu \frac{k^2}{\overline{\epsilon}}, \quad (5a)$$

$$\frac{\partial k}{\partial t} + \bar{u} \frac{\partial k}{\partial x_1} + \bar{v} \frac{\partial k}{\partial x_2} = \nu_T \left( \frac{\partial \bar{u}}{\partial x_2} \right)^2 - \varepsilon + \frac{\partial}{\partial x_2} \left[ \left( \nu + \frac{\nu_T}{\sigma_k} \right) \frac{\partial k}{\partial x_2} \right], \quad (5b)$$

$$\frac{\partial \overline{\epsilon}}{\partial t} + \bar{u} \frac{\partial \overline{\epsilon}}{\partial x_1} + \bar{v} \frac{\partial \overline{\epsilon}}{\partial x_2} = C_{\varepsilon 1} f_1 \frac{\overline{\epsilon}}{k} \nu_T \left( \frac{\partial \bar{u}}{\partial x_2} \right)^2 - C_{\varepsilon 2} f_2 \frac{\overline{\epsilon}^2}{k} + E + \frac{\partial}{\partial x_2} \left[ \left( \nu + \frac{\nu_T}{\sigma_\varepsilon} \right) \frac{\partial \overline{\epsilon}}{\partial x_2} \right], \quad (5c)$$

see Ref. 44. Here, $k$ is the turbulent kinetic energy and $\overline{\epsilon}$ is the isotropic turbulent dissipation, i.e. the term that controls the dissipation rate of $k$. The isotropic dissipation (which is zero at the wall) is related to the dissipation $\epsilon$ by $\overline{\epsilon} = \varepsilon_0 + \overline{\epsilon}$, where $\varepsilon_0$ is the value of the turbulent dissipation at $x_2 = 0$. The system (5a)-(5c) contains several closure coefficients and empirical damping functions, which act directly on these coefficients. Without the damping functions the $k - \varepsilon$ model would not be able to provide accurate predictions in the viscous near-wall region [44]. The Launder-Sharma $k - \varepsilon$ model [19] is obtained by specifying these damping functions as follows

$$f_\mu = \exp \left[ -3.4 / (1 + Re_T / 50) \right], \quad f_1 = 1,$n

$$f_2 = 1 - 0.3 \exp \left[ -Re_T^2 \right], \quad \varepsilon_0 = 2\nu \left( \frac{\partial \sqrt{k}}{\partial x_2} \right)^2,$n

$$E = 2\nu \nu_T \left( \frac{\partial^2 \overline{u}}{\partial x_2^2} \right)^2, \quad (6)$$

where $Re_T \equiv k^2 / \overline{\epsilon} \nu$. In the case of the Launder-Sharma $k - \varepsilon$ model, the closure coefficients have the following values

$$C_\mu = 0.09, \quad C_{\varepsilon 1} = 1.44, \quad C_{\varepsilon 2} = 1.92, \quad \sigma_k = 1.0, \quad \sigma_\varepsilon = 1.3. \quad (7)$$

We do not expect these values to be generally applicable 'best' values, and other $k - \varepsilon$ models do use different values. For instance, the Jones-Launder model [14], which only differs from (6) by a slightly different $f_\mu$, uses

$$C_\mu = 0.09, \quad C_{\varepsilon 1} = 1.55, \quad C_{\varepsilon 2} = 2.0, \quad \sigma_k = 1.0, \quad \sigma_\varepsilon = 1.3. \quad (8)$$

We refer to Wilcox [44] for further discussion on $k - \varepsilon$ type models and their limitations.
3.1. Classical identification of closure coefficients

The values of the closure coefficients in (7) are classically chosen by reference to fundamental flow problems. We illustrate how the nature of the coefficients leads to some ambiguity regarding their values, and how flow independent single best values are unlikely to exist. One such a fundamental flow problem often considered is homogeneous, isotropic, decaying turbulence. In this case the $k$ and $\varepsilon$ equations (4a)-(5c) (without damping functions) simplify to

$$\frac{dk}{dt} = -\varepsilon, \quad (9)$$

$$\frac{d\varepsilon}{dt} = -C_\varepsilon \varepsilon^2 \frac{k}{k}. \quad (10)$$

These equations can be solved analytically to give

$$k(t) = k_0 \left( \frac{t}{t_0} \right)^{-n}, \quad (11)$$

with reference time $t_0 = nk_0/\varepsilon_0$ and $n = 1/(C_\varepsilon - 1)$. And thus,

$$C_\varepsilon = \frac{n + 1}{n}. \quad (12)$$

The standard value for $n$ is such that $C_\varepsilon = 1.92$. However, this is by no means a hard requirement and other models do use different values for $C_\varepsilon$. For instance, the RNG $k - \varepsilon$ model uses a modified $C_\varepsilon = 1.68$ and the $k - \tau$ model (essentially a $k - \varepsilon$ model rewritten in terms of $\tau = k/\varepsilon$ [38]) uses $C_\varepsilon = 1.83$ [44]. Also, the experimental data from Ref. 24 suggests that most data agrees with $n = 1.3$, which corresponds to $C_\varepsilon = 1.77$.

The coefficient $C_\mu$ is calibrated by considering the approximate balance between production and dissipation which occurs in free shear flows, or in the inertial part of turbulent boundary layers. This balance can be expressed as

$$P = \nu_t \left( \frac{\partial \bar{u}_1}{\partial x_2} \right)^2 = C_\mu \frac{k^2}{\varepsilon} \left( \frac{\partial \bar{u}_1}{\partial x_2} \right)^2 = \varepsilon. \quad (13)$$

Equation (13), can be manipulated together with the turbulent-viscosity hypothesis $-\bar{u}_1' \bar{u}_2' = \nu_t \partial \bar{u}_1/\partial x_2$ to yield $-\bar{u}_1' \bar{u}_2' = \varepsilon (\partial \bar{u}/\partial x_2)^{-1}$, which in turn yields

$$C_\mu = \left( \frac{\bar{u}_1' \bar{u}_2'}{k} \right)^2. \quad (14)$$

The DNS data from Ref. 16 can be used to show that $\bar{u}_1' \bar{u}_2' \approx -0.30k$ (except close to the wall), such that $C_\mu = 0.09$ is the recommended value. Again however, different models use different values for $C_\mu$, such as $C_\mu \approx 0.085$ in the case of the RNG $k - \varepsilon$ model.
Another fundamental flow to be considered is fully developed (so \( Dk/Dt = De/Dt = 0 \)) channel flow. The resulting simplified governing equations allows us to find the following constraint amongst several parameters \[31\]

\[
\kappa^2 = \sigma_\varepsilon C^{1/2}_\mu (C_{\varepsilon 2} - C_{\varepsilon 1}),
\]

where \( \kappa \) is the von-Karman constant. It should be noted that the suggested values (7) satisfy this constraint only approximately. Using (7) in (15) gives \( \kappa \approx 0.43 \) instead of the 'standard' value of 0.41.

The following constraint (between \( C_{\varepsilon 1} \) and \( C_{\varepsilon 2} \)) can be found by manipulating the governing equations of uniform (i.e. \( \partial \bar{u}_1/\partial x_2 = \text{constant} \)) shear flows \[31\]

\[
\left( \frac{\mathcal{P}}{\varepsilon} \right) = \frac{C_{\varepsilon 2} - 1}{C_{\varepsilon 1} - 1},
\]

where the non-dimensional parameter \( \mathcal{P}/\varepsilon \) is the ratio between the turbulent production \( \mathcal{P} \) and dissipation \( \varepsilon \). Tavoulakis et al. \[41\] measured \( \mathcal{P}/\varepsilon \) for several uniform shear flows. They reported values between 1.33 and 1.75, with a mean around 1.47. Note however, that (16) becomes 2.09 with the standard values for \( C_{\varepsilon 1} \) and \( C_{\varepsilon 2} \), which is significantly different from the mentioned experimental values.

The parameter \( \sigma_k \) can be considered as a turbulent Prandtl number, defined as the ratio of the momentum eddy diffusivity and the heat-transfer eddy diffusivity. These quantities are usually close to unity, which is why the standard value for \( \sigma_k \) is assumed to be 1.0. As noted in Ref. 30, no experimental data can be found to justify this assumption. And again, we see a range of recommended values amongst the different variations of the \( k-\varepsilon \) model. For instance, the RNG \( k-\varepsilon \) model uses \( \sigma_k = 0.72 \) \[44\].

The parameter \( \sigma_\varepsilon \) controls the diffusion rate of \( \varepsilon \), and its value can be determined by using the constraint (15), i.e.

\[
\sigma_\varepsilon = \frac{\kappa^2}{C^{1/2}_\mu (C_{\varepsilon 2} - C_{\varepsilon 1})}.
\]

Finally, it should be noted that the 'constant' value of the von Karman constant (0.41) is being questioned. An overview of experimentally determined values for \( \kappa \) is given in Ref. 46, which reports values of \( \kappa \) in \([0.33, 0.45]\)

3.2. Numerical solution of the \( k-\varepsilon \) model

To obtain efficient numerical solutions for the boundary-layer problem (4a)-(5c) we used the program EDDYBL of Ref. 43, which we modified slightly to make it more suitable for our purpose. EDDYBL is a two-dimensional (or axisymmetric), compressible (or incompressible) boundary-layer program for laminar, transitional and turbulent boundary layers. This program has evolved over three decades and is based on a code originally developed by Price and Harris in 1972 \[43\]. The advantage of using a boundary-layer approximation
rather than a full RANS code, is that a boundary-layer code allows for quicker numerical simulations, and thus avoids the need of a surrogate model.

Parabolic systems of equations such as the boundary-layer equations can, in general, be solved using unconditionally stable numerical methods. EDDYBL uses the variable-grid method of Blottner [1], which is a second-order accurate finite-difference scheme designed to solve the turbulent boundary-layer equations. This scheme uses a three-point forward-difference formula in the stream-wise direction, central differencing for the normal convection term and conservative differencing for the diffusion terms.

We verify that the discretization error is small enough such it does not dominate over the uncertainties we want to quantify. The rate at which the grid-point spacing increases in normal direction is set such that the first grid point satisfies $\Delta y^+ < 1$, which provides a good resolution in the viscous layer. Initially, the maximum number of points in the normal direction is set to 101, although EDDYBL is capable of adding more points if needed to account for boundary-layer growth. The maximum number of stream-wise steps is set high enough such that EDDYBL has no problems reaching the specified $s_{stop}$, i.e. the final arc length in stream-wise direction. Using this setup we verify that the discretization errors are substantially smaller than the uncertainties present in the model and data. To give an example of the magnitude of the discretization error, we computed the boundary layer over the curved airfoil-shaped surface of Ref. 34 with $s_{stop} = 20.0 \text{ [ft]}$ for both our standard mesh with the first grid point below $y^+ = 1$, and on a finer mesh with the first 15 points below $y^+ = 1$. The maximum relative error between the two predicted velocity profiles was roughly 0.3%, which is well below the expected variance in the model output that we might see due to for instance the uncertainty in the closure coefficients. Discretization error is assumed to be negligible hereafter.

4. Experimental data

EDDYBL comes with configuration files which mimic the experiments described in the 1968 AFOSR-IFP-Stanford conference proceedings [5]. From this data source, we selected one zero pressure-gradient flow, and 12 flows from other types of available pressure gradients, which range from favourable ($dp/dx < 0$) to strongly adverse ($dp/dx > 0$) gradients. These 13 flows are described in table 1. The identification number of each flow is copied from Ref. 5. According to Ref. 42, the flows are identified as being 'mildly adverse', 'moderately adverse' etc, based upon qualitative observations of the velocity profile shape with respect to the zero-pressure gradient case. We plotted the experimentally determined, non-dimensional, streamwise velocity profiles in Figure 1. As usual, the normalized streamwise velocity is defined as $u^+ \equiv \bar{u}_1/\sqrt{\tau_{w}/\rho}$, where $\tau_{w}$ is the wall-shear stress. The normalized distance to the wall, displayed on the horizontal axis of Figure 1, is $y^+ \equiv x\sqrt{\tau_{w}/\rho}/\nu$. Too much weight should not be given to the classifications of the severity of the adverse gradients, since some flows (such as 2400) experience multiple gradient types along the spanwise direction. Also, when we try to justify the classification based upon the velocity profile
shape we find some discrepancies. For instance, based upon the profile shape alone, we would not classify flow 1100 as mildly adverse, or 2400 as moderately adverse.

To obtain an estimate of the spread in closure coefficients, we calibrate the $k-\varepsilon$ model for each flow of table 1 separately, using one velocity profile as experimental data.

Use of experimental data in the viscous wall region is worthy of a separate discussion. On one hand, Reynolds stresses tend to zero when approaching the wall, so that calibrating the turbulence model using data from the first few wall units does not make sense; moreover, in the whole viscous layer the model is dominated by damping functions (6), not calibrated here, introduced to enforce asymptotic consistency as $y^+ \to 0$. As a consequence, little information is obtained from the measurements here. On the other hand, obtaining reliable measurements close to the wall can be difficult due to limited spatial resolution, see e.g. [13]. Therefore, most experimental datasets of [5] do not include points in this region. Additional difficulties may arise according to the experimental technique in use: for instance, outliers due to additional heat losses near the wall are not uncommon in hot-wire measurements, and special corrections are needed to fix the problem [18]. Numerical results from Direct Numerical Simulations (DNS) could be used instead of experimental data sets. Nevertheless, there is little DNS data with high enough values of the friction Reynolds number $Re_\tau$ to allow for a sufficiently extended logarithmic region (see [29] for a recent survey).

More generally, our goal is to introduce and test a methodology that can be applied to complex, high-$Re$ flows for which DNS is simply not feasible. The effect of excluding near-wall data from the calibration data set is investigated through numerical experiments presented in Section 6.2.

Figure 1: The experimental data from Ref. 5.
<table>
<thead>
<tr>
<th>Identification</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1400</td>
<td>Zero</td>
<td>Equilibrium boundary layer at constant pressure</td>
</tr>
<tr>
<td>1300</td>
<td>Fav</td>
<td>Near-equilibrium boundary layer in moderate negative pressure gradient</td>
</tr>
<tr>
<td>2700</td>
<td>Fav</td>
<td>Equilibrium boundary layer in mild negative pressure gradient</td>
</tr>
<tr>
<td>6300</td>
<td>Fav</td>
<td>Near-equilibrium boundary layer growing beneath potential flow on model spillway</td>
</tr>
<tr>
<td>1100</td>
<td>Mild adv</td>
<td>Boundary layer in diverging channel</td>
</tr>
<tr>
<td>2100</td>
<td>Mild adv</td>
<td>Boundary layer on large airfoil-like body; pressure gradient first mildly negative, then strongly positive, with eventual separation</td>
</tr>
<tr>
<td>2500</td>
<td>Mild adv</td>
<td>Equilibrium boundary layer in mild positive pressure gradient</td>
</tr>
<tr>
<td>2400</td>
<td>Mod adv</td>
<td>Initial equilibrium boundary layer in moderate positive pressure gradient; pressure gradient abruptly decreases to zero, and flow relaxes to new equilibrium</td>
</tr>
<tr>
<td>2600</td>
<td>Mod adv</td>
<td>Equilibrium boundary layer in moderate positive pressure gradient</td>
</tr>
<tr>
<td>3300</td>
<td>Mod adv</td>
<td>Boundary layer, initially at constant pressure, developing into equilibrium flow in moderate positive pressure gradient</td>
</tr>
<tr>
<td>0141</td>
<td>Str adv</td>
<td>Boundary-layer with strong adverse pressure gradient, source [17]</td>
</tr>
<tr>
<td>1200</td>
<td>Str adv</td>
<td>Boundary layer in diverging channel with eventual separation</td>
</tr>
<tr>
<td>4400</td>
<td>Str adv</td>
<td>Boundary layer in strong positive pressure gradient</td>
</tr>
</tbody>
</table>
5. Methodology: Calibration and Prediction

Our methodology consists of two major parts: calibration and prediction. In the calibration stage (Section 5.1) posterior distributions on closure coefficients are identified for each of a set of 13 boundary-layer flows. These posteriors are then summarized with Highest Posterior Density (HPD) intervals in Section 5.3. The results will give a first indication of the extent to which posterior distributions of turbulence closure coefficients $\theta$ are case-dependent. This stage can also be seen as a setup stage for the predictive mechanism of our methodology.

To make predictions of a new, unseen flow, we combine the 13 posterior distributions for $\theta$ using p-boxes (Section 5.4). These p-boxes encapsulate the effect of both the 13 individual posterior uncertainties (due to the data not exactly identifying a single optimal $\theta$, but rather a probability distribution different from a Dirac function), and the variability of $\theta$ between cases.

5.1. Calibration framework

Bayesian calibration requires selection of joint prior distribution for the calibration parameters and a joint pdf (or statistical model) describing the likelihood function.

In our turbulence model calibration we have a large number of accurate observations, and a belief that model inadequacy will dominate the error between reality and prediction. In this situation we expect the prior on closure coefficients to be substantially less influential than the joint pdf. We therefore impose uniform priors on closure coefficients, on intervals chosen to: (i) respect mild physical constraints, and (ii) ensure the solver converges in most cases.

After the calibration we verify that the posterior pdf is not unduly constrained by the prior intervals: if we find that one of the informed marginal posteriors is truncated, we simply re-perform the calibration with a wider prior range for the truncated coefficient. We also perform a posterior model checking, in the sense that we verify that sufficient overlapping between the posterior model distribution and the calibration data interval exists.

To specify the joint likelihood we start from the framework of Cheung et al. [3], who use a multiplicative model inadequacy term, modelled as a Gaussian process in the wall-normal direction. Multiplicative error models are less common than additive errors, but may be useful in many engineering situations: here, it allows to enforce automatically that the random velocity profiles satisfy a no-slip wall condition (see [3]). By considering multiple different flows we have additional modelling choices. Unlike Cheung et al., we choose to calibrate closure coefficients and model-inadequacy hyper-parameters independently for each flow, and examine the variability between flows in a post-calibration step.

Let the experimental observations from flow-case $k \in \{1, \cdots, N_C\}$ be $z_k = [z^1_k, z^2_k, \cdots, z^{N_k}_k]$. Here $N_k$ is the number of scalar observations in flow-case $k$, and $z^i_k$ is the scalar observation at location $y^+_k > 0$, where in the following we work in $y^+$-units. Following Ref. 3, we assume the observation noise $\lambda_k = [\lambda^1_k, \cdots, \lambda^{N_k}_k]$ is known and uncorrelated at all measurement points. Furthermore, the closure coefficients and flow parameters for case $k$ are denoted $\theta_k$ and $t_k$ respectively.
The flow parameters include specification of the pressure-gradient as a function of the $x$-coordinate. The observation locations $y^+_k$, noise $\lambda_k$, and flow parameters $t_k$ are modelled as precisely known explanatory variables. In the case that substantial uncertainties existed in the experiments these could be modelled stochastically as nuisance parameters.

A statistical model accounting for additive Gaussian noise in the observations **via an additive term** and model inadequacy **via a multiplicative term** is: $\forall k \in \{1, \cdots, NC\}$

$$z_k = \zeta_k(y^+_k) + e_k,$$  

(18a)

$$\zeta_k(y^+_k) = \eta_k(y^+_k) \cdot u^+(y^+_k, t_k; \theta_k),$$  

(18b)

where $u^+(\cdot, \cdot; \cdot)$ is the simulation code output, and the multiplication is applied element-wise to its arguments. Observational noise is modelled as

$$e_k \sim N(0, \Lambda_k), \quad \Lambda_k := \text{diag}(\lambda_k),$$

and the model-inadequacy term $\eta_k(\cdot)$ is a stochastic process in the wall-distance $y^+$ modelling the relative error between the code output and the true process. Therefore (18a) represents the difference between the true process $\zeta_k$ and the measurement observations, and (18b) the difference between $\zeta_k$ and model predictions. Together they relate $\theta_k$ to $z_k$.

Cheung et. al. consider three models of this form, which differ only in the modelling of $\eta$. They compared the posterior evidence, and showed that modelling $\eta$ as a correlated Gaussian process yielded by far highest evidence of the three models considered [3]. We therefore adopt the same strategy and model each $\eta_k$ as a Gaussian process with unit mean (dropping the subscript $k$ for convenience):

$$\eta \sim \text{GP}(1, c_\eta),$$  

(19)

and the simple, homogeneous covariance function

$$c_\eta(y^+, y'^+ \mid \gamma) := \sigma^2 \exp \left[ -\left( \frac{y^+ - y'^+}{10^\alpha l} \right)^2 \right],$$  

(20)

where $y^+$ and $y'^+$ represent two different measurement points along the velocity profile, and $l$ is a user-specified length scale. We fix this length scale to 5.0, which is the $y^+$ value that denotes the end of the viscous wall region. The smoothness of the model-inadequacy term is controlled by the correlation-length parameter $\alpha$, and its magnitude by $\sigma$. Both $\alpha$ and $\sigma$ are unknown, and must be obtained via calibration from the data, and form a hyper-parameter vector $\gamma := [\alpha, \sigma]$.

A more boundary-layer specific model than (19), is described in Ref. 28. It attempts to account for the multi-scale structure of the boundary layer by allowing the correlation length to vary in $y^+$ direction. Together (18b) and (19) imply the relative model inadequacy $\sigma$ is independent of $y^+$, and that the correlation length is the same throughout the boundary layer. This model may be generalized to multiple dimensions by replacing $\eta(\cdot)$ with a multi-dimensional

A consequence of the above modelling choices is that the true process $\zeta$ is also modelled as a Gaussian process:

$$
\zeta \mid \theta, \gamma \sim \text{GP}(\mu_\zeta, c_\zeta) \tag{21}
$$

$$
\mu_\zeta(y^+ \mid \theta) = u^+(y^+, t; \theta)
$$

$$
c_\zeta(y^+, y'^+ \mid \theta, \gamma) = u^+(y^+, t; \theta) \cdot c_\eta(y^+, y'^+ \mid \gamma) \cdot u^+(y'^+, t; \theta),
$$

which is still centered around the code output. The assumption of normality is made mainly for convenience, and more general forms are possible. See Section 5.5 for a discussion on the choice of statistical model.

The likelihood evaluated at the measurement locations $y^{+,i}$ can now be written for each flow case $k$ independently as:

$$
p(z \mid \theta, \gamma) = \frac{1}{\sqrt{(2\pi)^N|K|}} \exp \left[ -\frac{1}{2} d^T K^{-1} d \right],
$$

$$
d := z - u^+(y^+)
$$

$$
K := \Lambda + K_\zeta.
$$

where

$$
[K_\zeta]_{ij} := c_\zeta(y^{+,i}, y^{+,j} \mid \theta, \gamma).
$$

Since in general the computational grid does not coincide with measurement locations we linearly interpolate the code output at $y^{+,i}$ where needed.

Note that the statistical model is based only on velocity data, and does not include constraints on other physical quantities, like e.g. the Reynolds stresses or the turbulent kinetic energy $k$. This choice is consistent with the fact that our version of $k - \epsilon$ is not realizable, and as such it does not contain any modelling assumption to prevent the normal Reynolds stresses from becoming negative, but only enforces them to satisfy physical constraints through the application of limiters. If we were to calibrate realizable $k - \epsilon$ we would try to preserve the realizability conditions. This would require adding constraints to the likelihood function so that zero probability is assigned to parameter combinations leading to unrealizable turbulence states.

5.2. Priors for $\theta$ and $\gamma$

Unlike Cheung et. al., we do not treat all closure coefficients as independent random variables in the prior. Instead we use the physical relations described
in Section 3.1 to constrain the value of two closure coefficients. Specifically we fix \( C_{\epsilon^1} \), by rewriting (16) as

\[
C_{\epsilon^1} = \frac{C_{\epsilon^2}}{P/\varepsilon} + \frac{P/\varepsilon - 1}{P/\varepsilon},
\]

where, similar to Ref. 30, we fix the ratio \( P/\varepsilon \) to 2.09. In our results, this choice locates the mode of the posterior for \( C_{\epsilon^2} \) relatively close to the standard value of 1.92. If we instead would have used a different (experimentally determined) value of \( P/\varepsilon \), the mode \( C_{\epsilon^2} \) would be located elsewhere. Whether or not our choice is reasonable has to be determined by the ability of the posterior distributions to capture the observed data, as outlined in Section 6.3. Two other possibilities we do not employ are: (i) to move \( P/\varepsilon \) into \( \theta \) and calibrate it along with the other parameters with some suitable prior, or (ii) model \( P/\varepsilon \) as an aleatory uncertainty, using the \( P/\varepsilon \) data from Ref. 41 to construct an approximate pdf \( p(P/\varepsilon) \). Also, we fix \( \sigma_\varepsilon \) using (15). Such a choice avoids running the boundary-layer code with non-physical parameter combinations.

All priors, for both the closure coefficients \( \theta \) and hyper-parameters \( \gamma \), are independent uniform distributions. The choice of interval end-points was made based on three factors: the spread of coefficients recommended in the literature, the range of coefficients for which the solver was stable, and avoidance of apparent truncation of the posterior at the edge of the prior domain. The range we used is specified in Table 2. We chose uniform distributions because we lack confidence in more informative priors for these parameters. We note however that some reasonable, informative priors can be obtained using the classical framework for coefficient identification (c.f. Section 3.1) in combination with multiple experimental measurements from different sources [30].

To obtain samples from the posterior distributions \( p(\theta \mid z) \), we employed the Markov-chain Monte Carlo (McMC) method [11]. We subsequently approximated the marginal pdf of each closure coefficient using kernel-density estimation, using the last 5,000 (out of a total of 40,000) samples from the Markov chain. It was observed that at 35,000 samples, the Markov chain was in a state of statistical convergence.

<table>
<thead>
<tr>
<th>coefficient ( \text{left boundary} )</th>
<th>right boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_{\epsilon^2} )</td>
<td>1.15 (-40%)</td>
</tr>
<tr>
<td>( C_\mu )</td>
<td>0.054 (-40 %)</td>
</tr>
<tr>
<td>( \sigma_k )</td>
<td>0.450 (-45 %)</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.287 (-30 %)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.0</td>
</tr>
<tr>
<td>( \log \alpha )</td>
<td>0.0</td>
</tr>
</tbody>
</table>
5.3. Summarizing posteriors: HPD intervals

The methodology of Section 5.1 will be applied to 13 flow cases, resulting in 13 posteriors on $[\theta, \gamma]$. The large amount of information (see e.g. Figure 2 and 3 in the results section) is difficult to visualize. In other words the posteriors must be summarized, and we do this with intervals. In the remainder we make the assumption that closure coefficients are approximately independent and uncorrelated. This is justified by Figure 3. This assumption allows us to work with 1d marginal pdfs of the coefficients, rather than the full multi-dimensional posteriors.

We compute Highest Posterior Density (HPD) intervals on 1d marginals to summarize the posteriors. An HPD interval is a Bayesian credible interval which satisfies two main properties, namely:

1. The density for every point inside the interval is greater than that for every point outside the interval.
2. For a given probability content $1 - \beta$, $\beta \in (0, 1)$, the interval is of the shortest length.

We use the algorithm of Chen et. al. [2] to approximate the HPD intervals using the obtained MCMC samples. To do so, we first sort the samples of the $Q$ closure coefficients $\theta^q, q = 1, 2, \cdots, Q$ in ascending order. Then, if we let $\{\theta^q_j, j = 1, 2, \cdots, J\}$ be the MCMC samples from $p(\theta^q | z)$, the algorithm basically consists of computing all the $1 - \beta$ credible intervals and selecting the one with the smallest width. For a given $j$, we can use the empirical cumulative-distribution function to approximate the $1 - \beta$ interval by computing the first $\theta^q_s$ which satisfies the inequality

$$\sum_{i=j}^J \mathbb{1}_{\theta^q_i \leq \theta^q_s} \geq [J (1 - \beta)],$$

where $\mathbb{1}_{\theta^q \leq \theta^q_s}$ is the indicator function for $\theta^q_i \leq \theta^q_s$ and $[J (1 - \beta)]$ is the integer part of $J (1 - \beta)$. Secondly, if we let $\theta^q_{(i)}$ be the smallest of a set $\{\theta^q_i\}$, then the first $\theta^q_s$ for which (24) is satisfied simply is $\theta^q_{(J (1 - \beta))}$. Thus, the $j^{th}$ credible interval is given by $\theta^q_{(J + [J (1 - \beta)])} - \theta^q_{(j)}$, and the HPD interval for $\theta^q$ is found by solving

$$\min_j \theta^q_{(J + [J (1 - \beta)])} - \theta^q_{(j)}, \quad 1 \leq j \leq J - [J (1 - \beta)].$$

The algorithm of Chen assumes a uni-modal posterior pdf, although it could possibly be extended to deal with multi-modal pdfs [2]. This assumption is shown to be justified in our case in Section 6.5.

5.4. Predictive framework: P-boxes

So far we have only discussed identification of $\theta$ for flows with data. The final purpose of this work is to establish uncertainties on $k-\varepsilon$ model predictions.
under flow conditions \( t^* \) at which no measurements are available. To achieve this we must assess the effect of all sources of uncertainty on the model prediction at \( t^* \). We require a method that can simultaneously represent solution variability within- and between the posteriors \( p(\theta, \gamma \mid z_k, t_k), k = 1, 2, \cdots, N_C \), where \( N_C = 13 \) in this work.

Our approach is to construct a probability box (p-box) for the output of the model at the new condition \( t^* \), using coefficients sampled from the HPD intervals of each of the 13 cases. A p-box is commonly used to visualize possible outcomes due to a combination of epistemic and aleatory uncertainty [26]. Examples of p-boxes are shown in Figure 13. In our case - roughly speaking - the slant of the box represents the width of individual HPD intervals, and the width of the box the variability between HPD intervals obtained from calibrations against different data sets. More precisely they can be interpreted as bounds on the \( \zeta \) (or \( u^+ \)) value at any particular probability level, and therefore can be used to construct confidence intervals on the true process \( \zeta \) or code output \( u^+ \).

We define our p-box as follows: For flow case \( k \), let \( \Theta_50^k \) be a uniformly distributed random variable on the box given by the 50% HPD intervals of the posterior for \( \theta \). Propagate this variable through the flow-code for \( u^+ \) to obtain

\[
Z_k(y^+_i) := u^+(y^+_i, t^*; \Theta_50^k),
\]

a random variable representing the effect of posterior uncertainty in case \( k \) on the model output at conditions \( t^* \). Here \( y^+_i \) is the location at which an uncertainty estimate is required, which need not correspond to a sample location in any of the calibrated flow cases. Let \( F_k(u^+) = P(Z_k \leq u^+) \) be the cumulative density function of \( Z_k \). Then the p-box \( \mathbb{D} \) is defined by

\[
\mathbb{D}(u^+) := \{ r \in [0, 1] \mid F(u^+) \leq r \leq F(u^+) \},
\]

\[
\overline{F}(u^+) := \min_{k \in \{1, \ldots, 13\}} F_k(u^+),
\]

\[
\underline{F}(u^+) := \max_{k \in \{1, \ldots, 13\}} F_k(u^+)
\]

i.e. the envelope formed by this collection of \( k \) cdfs. To construct a worst-case 90% confidence interval we find the end-points \( u^{+\text{L}} \) and \( u^{+\text{U}} \) such that

\[
\underline{F}(u^{+\text{L}}) = 0.05,
\]

\[
\overline{F}(u^{+\text{U}}) = 0.95.
\]

The interval \( I = [u^{+\text{L}}, u^{+\text{U}}] \) is our final estimate of solution uncertainty due to modelling inadequacy in the \( k - \varepsilon \) model for \( u^+(y^+_i) \) at conditions \( t^* \).

To construct the p-box in (26) numerically we use empirical cdfs:

\[
F_k(u^+) = P(Z_k \leq u^+) \approx \frac{1}{S} \sum_{j=1}^{S} \mathbb{1}_{u_j^+ \leq u^+},
\]

(27)

where \( u_j^+ \) are \( S \) samples from \( Z_k \) obtained using Monte-Carlo. An approximation to \( \mathbb{D} \) is then readily calculated.
Note that because $Z_k$ is based only on the flow-code output and not on the true process $\zeta$, the effect of the model inadequacy term $\eta$ is not included in our estimates. If a variable other than $u^+$ were of interest, we could define the p-box in exactly the same way (using skin-friction $C_f$ as an example):

$$ Y_k(x) := C_f(x, t^*; \Theta_50^k). $$

That this is possible is a consequence of representing model inadequacy via closure coefficients. It is not possible if we base estimates on $\eta$-like terms. For the validity of the confidence intervals we are relying on the uncertainties in $\theta$ accounting for the majority of model error. We acknowledge that the choice of $50\%$ HPD intervals plays a role in the p-box size, and this tuning parameter could be eliminated by replacing $\Theta_50^k$ and $\Gamma_50^k$ by the posteriors for case $k$, at the cost of increasing the size of the p-boxes.

5.5. Discussion

In the above we are attempting to capture model error in two different ways:

1. Via the traditional (Kennedy and O’Hagan) Gaussian process (GP) “model discrepancy” term, $\eta(\cdot)$, and
2. Via representing the variability of $\theta$ across flows (using p-boxes).

In doing so we have introduced some redundancy, since model errors can be captured by either one of these two methods, or a combination of them both. Our aim is to capture the majority of model error via $\theta$-variability, for the following reasons.

The Gaussian process $\eta(\cdot)$ has a very general form, able to capture a large space of smooth model-data error as a function of $y^+$. It has no physical content (other than the assumption of smoothness). Its form is geometry and flow-topology dependent: e.g. in our case the calibrated $\eta(\cdot)$ can not be used to predict error in any other variable than $u^+$, in any other geometry. The natural generalization to more complex flows is a multi-dimensional Gaussian process, which would necessitate estimation of a large number of hyper-parameters.

In contrast, varying $\theta$ does not allow for general velocity profiles. Only those profiles that satisfy the governing equations (for some value of $\theta$) are represented, hence certain constraints based on physical modelling are automatically satisfied. Since each component of $\theta$ corresponds to some empirical modelling assumption, one might also see $\theta$ as an (incomplete) parameterization of modelling assumptions. Furthermore, irrespective of flow topology, uncertainties on $\theta$ can be propagated through the simulation code to estimate model error in the output. That is, distributions of $\theta$ derived from calibration on a class of boundary-layer flows, can be applied to estimate model error on any flow. The accuracy will depend on the extent to which boundary-layer modelling error dominates in the new flow, e.g. the error in fully attached subsonic flows over aerofoils might be successfully judged.

In short we consider representing model error in model-coefficient space has several advantages, compared to representing it in the data space. However all
error can not be represented in $\theta$-space – e.g. error due to model form is not accounted for. Therefore, to characterize the real relationship between data and model output in the statistical model, and to prevent over-fitting, $\eta$ is necessary in (18b). Of course any choice of statistical model represents a modelling assumption, and in Section 6.6 we examine the sensitivity of $\theta$-posteriors to the choice of $\eta$.

6. Results and discussion

6.1. Marginal posterior pdfs

Calibration of the $k-\varepsilon$ model, using the experimental data described in Section 4, and the statistical model of Section 5 was performed. The marginal posterior pdfs of all four parameters in $\theta$, for all of the 13 calibration cases are shown in Figure 2. There we see that the data has been informative for $C_{\varepsilon^2}$ in all cases, resulting in sharply peaked posteriors. This is in contrast to $C_{\mu}$, which has been only weakly informed. Experience suggests that for these cases the level of informativeness is predicted by the sensitivity of the data $u^{+}(\theta)$ to the individual parameters. Parameters with the largest sensitivities are the best identified, see section 6.4. The calibrations have provided us with a bit more information regarding $\sigma_k$, since more posterior distributions show clear modes compared to the $C_{\mu}$ results. The spread of coefficients between cases is also visible. For instance $C_{\varepsilon^2}$ values cluster around the center of the prior interval, while $\kappa$ is sharply identified at distinctly different values for the different flow conditions.

To examine dependencies between variables, we plot two-variable marginal posteriors of all pairs of closure coefficients, see Figure 3. This figure is constructed from the McMC traces for flow 3300, and is typical of the other flow cases which are not shown. Any correlation between two coefficients will be visible in such a plot. As can be seen, there is a weak negative correlation between $C_{\varepsilon^2}$ and $C_{\mu}$, and also between $C_{\varepsilon^2}$ and $\sigma_k$. However, overall the coefficients appear approximately independent, as they were in the prior. This observed independence may not be a coincidence, it may be a result of the design of the turbulence model in which the individual coefficients parameterize separate modelling assumptions.

6.2. $y^{+}$-cutoff sensitivity

In Section 4 we mention that we omit data in the viscous-wall region. To examine if our posterior distributions are critically affected by not including near-wall data, we perform a simple sensitivity analysis using flow case 1400. Originally, this flow contains data points from $y^{+} = 38.1$ onward. We add 10 artificial data points to $z$ between this value and $y^{+} = 1$. Next we perform 10 separate calibrations where each time we shift the lower bound of $z$ one point closer to the original lower bound of $y^{+} = 38.1$. We visualize the results in Figure 4 by plotting the HPD intervals for $C_{\varepsilon^2}$ and $\kappa$ versus $y^{+}_{1}$, i.e. the $y^{+}$ value corresponding to the first data point. From this figure it becomes clear
Figure 2: The marginal posterior distributions of the coefficients for the 13 cases of Table 1.
that the inclusion of near-wall data does not significantly alter the posteriors in this case. The other parameters in both $\theta$ and $\gamma$ show similar behaviour.

The sparsity of available experimental near-wall data, together with the indication that the posterior distributions do not change significantly justifies our exclusion of near-wall data.

6.3. Posterior model check

In Bayesian analysis it is good practice to assess the fit of the chosen model. We expect all observed data used in the calibration for each flow case to lie within the range of the posterior predictive distribution of the true process $\zeta$ for that case. It should be noted that this is not the same as validating the model, since it only ensures that the chosen model is capable of reproducing the observed data. It does not guarantee that it can also be used for predictions. In our model the variability in $\zeta$ can be broken down into that due to the explicit model inadequacy term $\eta(\gamma^+)$, and that due to uncertainty in $\theta$. The former can be obtained directly from (19) and the calibrated values of $\gamma = [\sigma, \alpha]$. The latter is just the posterior of $u^+(\theta)$, and can be computed using the $u^+$ traces stored during the McMC calibration run. This is equivalent to propagating posterior samples of $\theta$ through the $k - \varepsilon$ model as in Monte-Carlo.

In Figure 5 we show only the uncertainty due to $\theta$ for two flows. The posterior prediction for $u^+$ encompasses all the experimental data, and this is true for all the flows described in Table 1, even those in which a large $\eta$ was predicted. They are therefore all consistent (in the sense of Ref. 22: existence of an overlap between the predictions and the region of experimental uncertainty). In addi-

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig3}
\caption{A two-dimensional contour plot of the posterior $\theta$ samples from flow 3300.}
\end{figure}
Figure 4: The HPD intervals corresponding to calibrations with different $y^+$ cutoff values.
Figure 5: The mean and 3 standard deviations of posterior $u^+ (y^+, \theta)$ samples of a favourable, and a strongly adverse flow (bottom). The green line indicates the solution of the $k-\varepsilon$ model using the standard values (7), and the red dots represent the experimental data with error bars.

To illustrate the effect of also using $\eta$, we compare posterior distributions of $u^+$ and $\zeta$ in Figure 6. The mean of both distributions is the same, which could be inferred from (21). Thus, including a model inadequacy term of the form (19) results in a posterior distribution of the true process with the same mean as the posterior $u^+$ distribution, but a larger variance. When making predictions with the model this contribution to the variance should be included.

It is instructive to examine the effect of the coefficient uncertainties on quantities that are not directly measured. To that end we show the standard-deviation cloud for the normalized turbulent kinetic energy $k$ in Figure 7. Since the condition $k > 0$ is enforced by limiters for all samples, the posterior prob-
Figure 6: The posterior distribution of $u^+$, and the posterior distribution of $\zeta$. 
ability of negative $k$ is guaranteed to be zero. However, since we do not have experimental data for the normalized turbulent kinetic energy, we cannot verify if its posterior distribution is consistent with the true process of $k$.

6.4. Sobol indices

Figure 2 shows that, for a given flow case $k$, there is significant variation in the amount of information contained in the posterior closure-coefficient distributions. In an attempt to elucidate this behaviour, we perform global sensitivity analysis on $u^+$ with respect to $\theta$. If the measurement data is very sensitive to a particular parameter or set of parameters in $\theta$, we expect the corresponding posterior distribution to be well informed and vice versa. In particular we use
Sobol’ indices $S_w$, defined as [36]

$$D = \text{Var}_W \{u^+\},$$
$$D_w = \text{Var}_w \{E_w(u^+|\theta_w)\},$$
$$S_w = D_w/D$$

where $w \subset \{1, \ldots, Q\} = W$ indexes the components of $\theta$, $w' = W \setminus w$, and $\text{Var}_w \{\cdot\}$ indicates variance taken over priors of $\theta_w$, etc. The indices $S_w$ satisfy $\sum_{\mathcal{P}(W)} S_w = 1$, where $\mathcal{P}(W)$ is the power set of $W$. A value close to unity for $S_i$ can be interpreted as the coefficient corresponding to $i \in W$ being responsible for most of the total variance in $u^+$ on its own (also without interaction effects with other parameters). A value close to zero indicates an uninfluential parameter. The $D_w$ are computed using a polynomial approximation to the response in the stochastic space [20], given which the expectation and variance can be evaluated analytically [39, 40].

Using the described setup, we calculate the main effect indices $S_i$ with $i = 1, 2, 3, 4$ corresponding to $\{C_{\varepsilon_2}, C_{\mu}, \sigma_k, \kappa\}$ for $u^+(y^+)$. The results are shown in Figure 8. The ranking from most sensitive parameter to least sensitive one for the velocity profile is $C_{\varepsilon_2}, \kappa, \sigma_k, C_{\mu}$. This is the same ranking that we get when we sort the coefficients from most informed posterior distribution to least informed one, see Figure 2. Thus, the very low sensitivity of our data to the value of $C_{\mu}$ is an explanation for the lack of information in the posterior $C_{\mu}$ distributions.

Figure 8: The Sobol indices $S_i$ for flow 1400, with the velocity profile as QoI. The horizontal axis represents the direction normal to the wall for the stream-wise location $s = 16.3[ft]$.

This sensitivity analysis can be used in experimental design. E.g. we could ask: would adding friction coefficients $C_f$ to the dataset lead to better informed distributions? Using the Sobol indices for $C_f$, we attempt to answer the question.
in the absence of $C_f$ data. The $S_i$ corresponding to the friction coefficient can be found in Figure 9. The influence of $C_\mu$ is still very low, therefore we would not expect to significantly improve its identification. This is consistent with the results from [3], where the SA turbulence model was calibrated using both velocity profiles and friction coefficients. Still, some of their posterior distributions were uninformative as well.

![Figure 9: The Sobol indices $S_i$ for flow 1400, with $C_f$ as QoI. The horizontal axis represents the stream-wise direction.](image)

In Figures (8)-(9) we also show $\sum S_i$, i.e. the sum of all displayed Sobol indices. This sum is very close, but not equal to 1. This indicates that for the considered range in the closure coefficients, the interaction effects are low, i.e. the Sobol indices corresponding a combination of closure coefficients are small.

6.5. Coefficient variability across test-cases

The HPD summaries of the $\theta$-posteriors are shown in Figure 10 for all test-cases. In these plots the ordering of the test-cases on the horizontal axis corresponds roughly to increasingly unfavorable pressure gradients. The spread of the posterior modes of $C_{\varepsilon 2}$ is quite concentrated, they all lie relatively close to the standard value of 1.92. The small width of the HPD intervals (compared to the prior range), indicates that the posterior distributions have been well-informed by the data. Also notice that a slight downward trend of the HPD intervals can be observed with increasing pressure gradient, suggesting some physical effect not present in the model.

The spread of $C_\mu$ is relatively small, with most distributions centered close to 0.06, consistently lower than the standard value (0.09). The only clear exception is flow 2600, which also showed some outlier behaviour for $C_{\varepsilon 2}$. The parameter $\sigma_k$ shows a larger spread, although in general values above the standard value of 1.0 are preferred.
Most individual pdfs of $\kappa$ are quite well informed, but the modes are spread roughly between 0.31 and 0.46. Previous studies have looked at the spread of the von Karman constant. An overview is given in Ref. 46, which reports values of $\kappa$ between $[0.33, 0.45]$, roughly similar to the spread that we have observed. The spread of the $\kappa$ HPD intervals in Figure 10 can be qualitatively explained by considering the deviation of the experimental velocity profiles of Figure 1 from the standard log law $1/\kappa \ln(y^+) + C$. As can be seen from Figure 1, from roughly $y^+ = 30$ the velocity profiles overlap onto the standard log law. However, around $y^+ = 200$ the first profiles start to deviate from this law. Qualitatively, the profiles which show a larger deviation from the log law, are also the ones which show a lower $\kappa$ HPD interval compared to the rest.

The HPD intervals of the hyper-parameters $\sigma$ and log $\alpha$ can be found in Figure 11. Most posterior modes of $\sigma$ are located near the bottom edge of the domain, indicating that $\eta$ is small, and the experimental data can be matched well with appropriate choice of $\theta$ alone. The posterior modes for log ($\alpha$) all lie between 2.5 and 3.5, indicating that the remaining model error is correlated over a large fraction of the boundary layer [3]. In other words, a smooth model inadequacy term is preferred.

Figure 11 shows three clear outliers to this trend, flows 2400, 2500 and 1200. High $\sigma$ values in these cases indicate that the data can not be matched with $\theta$ alone.

6.6. Statistical model sensitivity

As mentioned in Section 5.5, the choice of $\eta(\cdot)$ represents a modelling assumption. We acknowledge that our choice of $\eta(\cdot)$ is not the best possible one in terms of physical modelling, e.g. the $\eta(\cdot)$ of [28] includes more physics. However, our goal is to capture model inadequacy mainly through $\theta$, while relying on $\eta$ to capture the remaining error. Hence, it is important to examine the sensitivity of the marginal $\theta$ posteriors to the form of $\eta$. Therefore, we re-computed all calibrations with $\eta = 1$. In Figure 12 we show the HPD intervals for both $\eta = 1$ and the original $\eta(\cdot)$ of (19). Notice that the HPD intervals of $\eta = 1$ are smaller, which was to be expected since no uncertainty due to model inadequacy is added. More importantly though, the spread of closure coefficients across flow cases is roughly the same for both statistical models. This indicates that it is indeed possible to represent the bulk of the model inadequacy in $\theta$ space, since choosing a radically different $\eta(\cdot)$ does not seem lead to a significantly different spread of $\theta$.

6.7. Prediction with uncertainties

The calibrations described above contain quantitative information on the accuracy of the $k - \varepsilon$ model for flat-plate flows. We apply the methodology described in Section 5.4 to estimate $u^+$ and the associated uncertainty for a new flat-plate flow not in the calibration set. We use the data (used only as ground-truth) from Ref. 25, which is boundary-layer data on a cylinder in axially symmetric flow. This is flow 3600 from the 1968 AFOSR-IFP-Stanford
Figure 10: The 50% HPD intervals of $\theta$ for the 13 cases of Table 1.
Figure 11: The 50 % HPD intervals of $\gamma$ for the 13 cases of Table 1.
Figure 12: The HPD intervals for $C_{e_2}$ and $\kappa$ for two statistical models.
conference [5]. The results for three $y^+$ stations ($y^+ = 46.2, 267.0$ and $1039.7$) are given in Figure 13, together with boxes representing the experimental data ±3σ.

The width of the p-boxes clearly dominates their slant, suggesting that case-to-case variability dominates uncertainty in individual posteriors. Therefore adding more experimental data to the 13 calibrations will not much reduce our overall uncertainty. Had we not included the strongly-adverse pressure-gradient cases, the width of the p-box would be significantly smaller. However this would simply correspond to a more restricted range of flows for which the analysis is valid. To tighten these error estimates, the class of considered flows must be restricted to flows in some sense similar to the new flow. It is not presently clear how to do this automatically. We note that our estimates of simulation error are larger than estimates of experimental error, as expected.

Finally we extract confidence intervals from the p-boxes of Figure 13. In Figure 14 we show 90% confidence intervals on $u^+$ values for case 3600, and compare them with the measurement data as a reference. Note that all intervals extracted from the p-boxes are consistent with the experimental data – indicating that modelling inadequacy has been successfully bounded for this case. It is clearly a tight bound on the low-$u^+$ side. That the upper bounds are not tight in some locations, is an indication that uncertainty in the closure coefficients remains substantial. To put this in perspective, consider that, for reasonable values of the coefficients, the $k-\epsilon$ model can produce solutions anywhere within our confidence intervals. It is expected that when applying this procedure, better turbulence models will result in tighter bounds. This is the subject of future work.

7. Conclusion

Using a Bayesian framework, we performed 13 separate calibrations of the closure coefficients in the standard $k-\epsilon$ model. The experimental data on which we calibrated consisted of velocity profiles from 13 boundary-layer flows, each subject to a different pressure gradient. This allowed us to investigate the resulting spread of the posterior coefficient distributions, caused by the range of considered pressure gradients. To summarize the spread we perform a Highest Posterior Density (HPD) analysis on all posterior distributions, which gives us 13 credible intervals of most-likely values for both the closure coefficients and the hyper-parameters used to parameterize the model-inadequacy term.

The results show a significant variation of coefficient-posteriors across this (very limited) range of flows, for $C_{\epsilon 2}$ and especially for $\kappa$. Other coefficients were not identified well enough by the data to allow for clear conclusions about their variability. In any case it is clear that a single calibration does not provide us with reliable coefficients, or an estimate of model error. On the other hand, we use all 13 posteriors to build an error estimate containing both between-case coefficient variability, and individual posterior uncertainty. This estimate is a probability-box (p-box) for model output, from which confidence intervals can be constructed. We demonstrate the methodology for a 14th flat-plate flow not
(a) The $u^+$ p-box for flow 3600 at $y^+ = 46.2$.

(b) The $u^+$ p-box for flow 3600 at $y^+ = 267.0$.

(c) The $u^+$ p-box for flow 3600 at $y^+ = 1039.7$.

Figure 13:
from the calibration set. The resulting confidence intervals include the reference measurements, indicating that the estimate is not too narrow.

Future work will involve attempting to narrow the confidence intervals further, by selection of a subset of cases from the calibration set that represent the new flow best. The current error estimate will be applied to more complex flows, to investigate its generality. Furthermore we will apply the analysis to other turbulence models, to determine if the conclusions hold for models that are generally considered more accurate than $k - \varepsilon$. The use of Bayesian model averaging to combine all the resulting information into a predictive framework is promising. Finally this methodology will be extended to design optimal experiments for identifying turbulence models.

Acknowledgements The present work was supported by the French 'Agence Nationale de la Recherche' (ANR) under contract ANR-11-MONU-008-002.


